KLASIFIKASI HALAMAN *WEB* MENGGUNAKAN METODE TF-IDF SVM PADA SISTEM TERDISTRIBUSI *MULTINODE CLUSTER*

oleh

Kwang Dharma Saelau

NIM : 612012016

Skripsi

Untuk melengkapi salah satu syarat untuk memperoleh

Gelar Sarjana Teknik

Program Studi Teknik Elektro

Fakultas Teknik Elektronika dan Komputer

Universitas Kristen Satya Wacana

Salatiga

Agustus 2016

PERPUSTAKAAN UNIVERSITAS
UNIVERSITAS KRISTEN SATYA WACANA
Jl. Diponegoro 52 – 60 Salatiga 50711
Jawa Tengah, Indonesia
Telp. 0298 – 321212, Fax. 0298 321433
Email: library@adm.uksw.edu ; http://library.uksw.edu

PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan di bawah ini:

Nama : KWANG DHARMA SAELAU

NIM : 612012016 Email : 612012016@student.uksw.edu

Fakultas : TEKNIK ELEKTRONIKA DAN KOMPUTER Program Studi : TEKNIK ELEKTRO

Judul tugas akhir : KLASIFIKASI HALAMAN WEB MENGGUNAKAN METODE TF-IDF SVM PADA SISTEM TERDISTRIBUSI MULTINODE CLUSTER

Pembimbing : 1. BANU W. YOHANES, M.Comp Sc.

2. HARTANTO K.W, M.T.

Dengan ini menyatakan bahwa:

1. Hasil karya yang saya serahkan ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar kesarjanaan baik di Universitas Kristen Satya Wacana maupun di institusi pendidikan lainnya.
2. Hasil karya saya ini bukan saduran/terjemahan melainkan merupakan gagasan, rumusan, dan hasil pelaksanaan penelitian/implementasi saya sendiri, tanpa bantuan pihak lain, kecuali arahan pembimbing akademik dan narasumber penelitian.
3. Hasil karya saya ini merupakan hasil revisi terakhir setelah diujikan yang telah diketahui dan disetujui oleh pembimbing.
4. Dalam karya saya ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali yang digunakan sebagai acuan dalam naskah dengan menyebutkan nama pengarang dan dicantumkan dalam daftar pustaka.

Pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari terbukti ada penyimpangan dan ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya saya ini, serta sanksi lain yang sesuai dengan ketentuan yang berlaku di Universitas Kristen Satya Wacana.

Salatiga, 4 JANUARI 2017

KWANG DHARMA SAELAU

F-LIB-080

PERPUSTAKAAN UNIVERSITAS
UNIVERSITAS KRISTEN SATYA WACANA
Jl. Diponegoro 52 – 60 Salatiga 50711
Jawa Tengah, Indonesia
Telp. 0298 – 321212, Fax. 0298 321433
Email: library@adm.uksw.edu ; http://library.uksw.edu

PERNYATAAN PERSETUJUAN AKSES

Saya yang bertanda tangan di bawah ini:

Nama : KWANG DHARMA SAELAU
NIM : 612012016 Email : 612012016@student.uksw.edu
Fakultas : FTEK Program Studi : TEKNIK ELEKTRO
Judul tugas akhir : KLASIFIKASI HALAMAN WEB MENGGUNAKAN METODE TF-IDF SVM PADA SISTEM TERDISTRIBUSI MULTINODE CLUSTER

Dengan ini saya menyerahkan hak *non-eksklusif** kepada Perpustakaan Universitas – Universitas Kristen Satya Wacana untuk menyimpan, mengatur akses serta melakukan pengelolaan terhadap karya saya ini dengan mengacu pada ketentuan akses tugas akhir elektronik sebagai berikut (beri tanda pada kotak yang sesuai):

☑ a. Saya mengijinkan karya tersebut diunggah ke dalam aplikasi Repositori Perpustakaan Universitas, dan/atau portal GARUDA

☐ b. Saya tidak mengijinkan karya tersebut diunggah ke dalam aplikasi Repositori Perpustakaan Universitas, dan/atau portal GARUDA**

* *Hak yang tidak terbatas hanya bagi satu pihak saja. Pengajar, peneliti, dan mahasiswa yang menyerahkan hak non-ekslusif kepada Repositori Perpustakaan Universitas saat mengumpulkan hasil karya mereka masih memiliki hak copyright atas karya tersebut.*

** *Hanya akan menampilkan halaman judul dan abstrak. Pilihan ini harus dilampiri dengan penjelasan/ alasan tertulis dari pembimbing TA dan diketahui oleh pimpinan fakultas (dekan/kaprodi).*

Demikian pernyataan ini saya buat dengan sebenarnya.

Salatiga, 5 JANUARI 2017

KWANG DHARMA SAELAU
Tanda tangan & nama terang mahasiswa

Mengetahui,

BANU W. YOHANES, M.CompSc.
Tanda tangan & nama terang pembimbing I

HARTANTO KW, M.T.
Tanda tangan & nama terang pembimbing II

PERNYATAAN BEBAS PLAGIAT

Saya, yang bertanda tangan di bawah ini:

NAMA : Kwang Dharma Saelau

NIM : 612012016

JUDUL : KLASIFIKASI HALAMAN *WEB* MENGGUNAKAN METODE TF-IDF SVM PADA SISTEM TERDISTRIBUSI *MULTINODE CLUSTER*

Menyatakan bahwa skripsi tersebut di atas bebas plagiat. Apabila ternyata ditemukan unsur plagiat di dalam skripsi saya, maka saya bersedia mendapatkan sanksi apa pun sesuai dengan aturan yang berlaku.

Salatiga, September 2016

Kwang Dharma Saelau

KLASIFIKASI HALAMAN *WEB* MENGGUNAKAN METODE TF-IDF SVM PADA SISTEM TERDISTRIBUSI *MULTINODE CLUSTER*

oleh

Kwang Dharma Saelau

NIM : 612012016

Skripsi ini telah diterima dan disahkan

Untuk melengkapi salah satu syarat memperoleh

Gelar Sarjana Teknik

dalam

Konsentrasi Teknik Telekomunikasi

Program Studi Teknik Elektro

Fakultas Teknik Elektronika dan Komputer

Universitas Kristen Satya Wacana

Salatiga

Disahkan oleh

Pembimbing I

Banu W. Yohanes, M.CompSc.

Tgl. 15-9-2016

Pembimbing II

Hartanto K.W, M.T.

Tgl. 14-9-2016

PERNYATAAN BEBAS PLAGIAT

Saya, yang bertanda tangan di bawah ini:

NAMA : Kwang Dharma Saelau

NIM : 612012016

JUDUL : KLASIFIKASI HALAMAN *WEB* MENGGUNAKAN METODE TF-IDF SVM PADA SISTEM TERDISTRIBUSI *MULTINODE CLUSTER*

Menyatakan bahwa skripsi tersebut di atas bebas plagiat. Apabila ternyata ditemukan unsur plagiat di dalam skripsi saya, maka saya bersedia mendapatkan sanksi apa pun sesuai dengan aturan yang berlaku.

Salatiga, September 2016

Kwang Dharma Saelau

INTISARI

Dengan semakin bertambah besarnya jumlah data dan halaman *web* di internet, semakin diperlukan pula sebuah *web crawler* yang memiliki efisiensi dan ketepatan dalam mencari sumber informasi yang diharapkan. Salah satunya adalah dengan memakai *focused web crawler*.

Dalam skripsi ini diusulkan sebuah *focused web crawler* yang diimplementasikan pada sebuah sistem terdistribusi Hadoop *Multinode Cluster* dengan metode TF-IDF SVM. Perancangan *focused web crawler* dengan sistem ini untuk mempercepat penelusuran halaman *web*, meningkatkan presisi dan akurasi topik serta meringankan proses kerja penelusuran.

Pada pengujian topik halaman *web* yang ditelusuri, digunakan 10 topik yaitu *smartphone*, banjir, bisnis, NBA(basket), olimpiade, pendidikan, pokemon, Portugal, sepak bola dan puasa. *Crawler* yang dirancang ini dapat menelusuri halaman *web* dengan kecepatan rata-rata 8,698 detik per halaman. Kemudian *crawler* ini juga memiliki *precision* 86,37% dan *recall* 66,68%.

Mengetahui,	Mengesahkan,	Penyusun,
Dr. Iwan Setyawan	Banu W. Yohanes, M.CompSc.	Kwang Dharma S.
Dekan	Pembimbing I	

ABSTRACT

With increasingly large amounts of data and web pages on the internet, the more necessary is also a web crawler that have high efficiency and accuracy in finding resources is expected. One of them is to use a focused web crawler.

In this thesis proposed a dedicated web crawler which is implemented in a distributed system Multinode Hadoop Cluster with TF-IDF SVM method. Design a dedicated web crawler with this system to speed up searches web pages, improving the precision and accuracy of topics as well as ease the work process search.

On testing the topic of the web pages searched, used 10 areas including smartphones, flooding, businesses, NBA (basketball), Olympic games, education, pokemon, Portugal, football and fasting. Crawler designed it can browse web pages with an average speed of 8,698 seconds per page. Then the crawler also has precision 86,37% and 66,68% recall.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas penyertaan-Nya selama ini, sehingga penulis dapat menyelesaikan perancangan serta penulisan tugas akhir skripsi ini sebagai syarat kelulusan di Fakultas Teknik Elektronika dan Komputer Universitas Kristen Satya Wacana.

Pada kesempatan ini penulis juga berterima kasih kepada berbagai pihak yang baik secara langsung maupun tidak langsung telah membantu penulis dalam menyelesaikan skripsi ini :

1. Tuhan Yesus yang selalu menyertai, membimbing, serta memberikan jalan terbaik sehingga penulis dapat menyelesaikan skripsi ini.
2. Kedua orang tua tercinta yang sudah membesarkan dan mendidik penulis dengan kasih sayang serta ajaran-ajaran baik yang sangat berarti.
3. Bapak Banu W. Yohanes, M.CompSc. dan Bapak Hartanto K.W, M.T. sebagai pembimbing I dan pembimbing II, terima kasih atas bimbingan dan saran yang telah diberikan kepada penulis selama mengerjakan skripsi ini.
4. Kakak penulis Kak Liang Arta, S.T. yang selalu mendukung penulis agar cepat menyelesaikan tugas akhir.
5. Staff pengajar dan Tata Usaha yang banyak memberikan pelajaran berharga.
6. Keluarga besar 2012 sebagai teman seperjuangan yang selalu memberi dukungan kepada penulis.
7. Teman-teman terdekat Adi, Samuel, Fandy, Keenan, Yohanes, Raynaldy, Vallicano, dan Feliks yang telah bersama-sama menghadapi masa perkuliahan.
8. Berbagai pihak yang tidak dapat dituliskan satu persatu, penulis ucapkan terima kasih.

Penulis menyadari skripsi ini masih banyak kekurangan dan jauh dari kesempurnaan. Maka dari itu melalui kata pengantar ini penulis sangat terbuka menerima kritik serta saran yang membangun sehingga penulis dapat memperbaiki kekurangan yang ada. Akhir kata penulis sangat berharap agar skripsi ini dapat memberikan manfaat dan kontribusi terhadap perkembangan teknologi di masa yang akan datang.

Salatiga, 25 Agustus 2016

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL